International Journal of Cardiology 224 (2016) 27–32

Contents lists available at ScienceDirect

International Journal of Cardiology

journal homepage: www.elsevier.com/locate/ijcard

Arrhythmia detection from heart rate variability by SDFA method

Raquel Romes Linhares

Faculty of Mathematics - UFU, Uberlândia, MG, Brazil

ARTICLE INFO

Article history:
Received 14 June 2016
Received in revised form 15 August 2016
Accepted 18 August 2016
Available online 24 August 2016

Keywords:
Arrhythmia detection
Smoothed detrended fluctuation analysis
Hurst estimator

ABSTRACT

The smoothed detrended fluctuation analysis (SDFA) based on DFA and the principal of wavelet shrinkage procedures is a scaling analysis method to represent the correlation properties of a time series. Since there is not a specific rule for the choice of the numbers of regressors in SDFA Method, we present here an asymptotic optimal choice. We carried out some Monte Carlo simulations on fractional Gaussian noise (FGN) models, to investigate the effect of the numbers of regressors in SDFA Method. We analyze the long dependence property in view of the SDFA method to compare 10 healthy and 10 unhealthy (with cardiac arrhythimia) RR time series randomly selected from databases of the PhysioBank. It is proposed that utilizing Hurst estimator by SDFA method, as an additional diagnostic tool may provide an indication of cardiac arrhythmia.

1. Introduction

The method of Detrended Fluctuation Analysis is proposed based on stochastic process theory and chaos dynamics, which can describe the long-range correlation of non-stationary time series. It has successfully been applied to different fields of interest, such as DNA sequences (see [16]), econophysics (see [6–8]), heart rate variability analysis (see [22]), long-time weather records (see [10]) and others (see [2,17]). Theoretical proof about DFA method can be found in Taqqu et al. [20], Bardet and Kammoun [1], Crato et al. [2] and Linhares [12].

A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. The nonparametric regression is the most important application of wavelet to the statistic, and it is based on the *principle of wavelet shrinkage*, which aims to reduce (and even remove) the noise present in a signal (see [3–5,21]). The wavelet transform splits the data into lowpass (approximation) portions and highpass (detail) portions. *Wavelet shrinkage* reduces the magnitude of terms in the highpass portions. Finally, the wavelet transform is inverted to get the denoised version of the data.

Linhares [13] propose the Smoothed Detrended Fluctuation Analysis method (SDFA), based on the Detrended Fluctuation Analysis and on the *wavelet shrinkage procedure*, where the SDFA method shows better performance. The procedure computes various statistic fluctuations measures $F(l)$, where l represents a window length. Get the *wavelet shrinkage estimator* $\hat{F}(l)$ of $F(l)$, for all $l\in\{4,5,\cdots,g(n)\}$. By varying the length l, the $\hat{F}(l)$ can be characterized by a scaling exponent, more precisely, the slope coefficient of the line obtained by the regression of $\hat{F}(l)$ on $\ln(l)$, with $l\in\{4,5,\cdots,g(n)\}$. The usual choice of $g(n)$ is $\lfloor \frac{n}{10} \rfloor$, where $\lfloor\cdot\rfloor$ indicates the integer part function and n is the length of the time series.

We propose here an asymptotic optimal choice for $g(n)$ and carried out some Monte Carlo simulations on fractional Gaussian noise (FGN) models, to investigate the effect of the numbers of regressors in SDFA Method.

Electrocardiography (ECG) is the process of recording the electrical activity of the heart over a period of time using electrodes placed on the skin. It is considered a representative signal of cardiac physiology, useful in diagnosing cardiac disorders. The time intervals between its various peaks, may contain useful information about the nature of disease afflicting the heart. Heart Rate Variability (HRV) is the physiological phenomenon of variation in the time interval between heart beats (see [19]). HRV is influenced by a number of physiological factors including gender, postural changes, ventilation and time of day. Pathological conditions such as congestive heart failure and diabetic neuropathy are associated with alterations in heart rate variability. Sammito and Böckelmann [18] present the main endogenous, exogenous and constitutional factors, and observed a decrease in HRV not only in connection with non-influenceable physiological factors such as age, gender and ethnic origin, but also in conjunction with a large number of acute and chronic diseases. Heart rate variability (HRV) data display nonstationary characteristics and exhibit long-range correlation (memory). The RR time series is the series of heartbeat interval, where R is a peak point respect to each heartbeat of the electrocardiography (ECG) wave, and RR is the interval between successive R. The discrete series of successive RR intervals (the tachogram) is the simplest signal that can be used to characterize heart rate variability (HRV) and has been applied in various clinical situations (see [11,15,19]). Cardiac arrhythmia is defined as a change in electrical activity within the heart which manifests as irregular heartbeats. In extreme cases, arrhythmia can be induced by damaged cardiac

http://dx.doi.org/10.1016/j.ijcard.2016.08.286
0167-5273/

tissue or abnormal cardiac anatomy which may lead to a stroke or heart attack. Current methods to detect arrhythmias include electrocardiogram (ECG) readings and analysis. Many studies propose the Hurst estimators, as an additional diagnostic tool may provide an indication of cardiac arrhythmia (see [15]).

In this work, we analyze the long dependence property in view of the SDFA method to compare 10 healthy and 10 unhealthy (with cardiac arrhythimia) RR time series from databases of the PhysioBank. It is proposed that utilizing Hurst estimator by SDFA method, as an additional diagnostic tool may provide an indication of cardiac arrhythmia.

The paper is organized as follows. Section 2 describes the Smoothed Detrended Fluctuation Analysis (SDFA) method and presents the proposed choice for $g(n)$ with statistical properties. The fractional Gaussian noise (FGN) processes and simulation study is presented in Section 3. In Section 4 we present the analysis of the 10 healthy and 10 unhealthy (with cardiac arrhythimia) RR time series. Section 5 gives the conclusions.

2. SDFA

Here we present the Smoothed Detrended Fluctuation Analysis that is based on the Detrended Fluctuation Analysis (see [16]) and on the wavelet shrinkage procedure (see [3–5]). The aim of this method is to reduce the noise present in the signal $\{F(l)\}_{l\in\{4,5,\cdots,g(n)\}}$, by reducing the magnitude of the wavelet coefficients, in order to obtain an efficient estimator for the Hurst exponent (H), that is widely used to quantify long range dependence in time series data.

To apply Smoothed Detrended Fluctuation Analysis method to a given time series $\{X_t\}_{t=1}^n$, it is necessary the following steps.

In a first step, a running sum of the observed variable $\{X_t\}_{t=1}^n$, is calculated

$$Y_t = \sum_{j=1}^{t} \left(X_j - \overline{X}\right),$$

for each $t\in\{1,2,\cdots,n\}$, where $\overline{X}$ is the average value of $\{X_t\}_{t=1}^n$. In the second step, we divide the time series $\{Y_t\}_{t=1}^n$ into $\lfloor \frac{n}{l} \rfloor$ nonoverlapping blocks, where each block has l observations. For each block, one fits a least-square line to the data. We denote by $\hat{Y}_t^l$, for $t=1,\cdots,n$, the adjusted fit for each t on each block of length l. After that, we detrend the time series $\{Y_t\}_{t=1}^n$, that is, in each block we calculate.

$$Z_t^l = Y_t - \hat{Y}_t^l, \quad \text{for all} \quad t\in\{1,\cdots,n\}. \tag{2.1}$$

In the third step, for each $l\in\{4,5,\cdots,g(n)\}$, we calculate the *root mean square fluctuation* of the new sequence,

$$\mathrm{F}(l) = \sqrt{\frac{1}{\tilde{n}} \sum_{t=1}^{\tilde{n}} \left(Z_t^l\right)^2}, \tag{2.2}$$

where $\tilde{n} = l \cdot \lfloor \frac{n}{l} \rfloor$.

Two functions are very important in the wavelet analysis: the mother and father wavelets. These wavelets generate a family of functions that can reconstruct a signal. A *mother wavelet* $\psi(\cdot)$ and a *father wavelet* (*or scale function*) $\phi(\cdot)$ are real functions $\psi,\phi:\mathbb{R}\to\mathbb{R}$ such that

$$\int_{\mathbb{R}} \psi(t)dt = 0, \qquad \int_{\mathbb{R}} \phi(t)dt = 1,$$

and satisfy some integrability conditions, that is, $\psi,\phi\in L^2(\mathbb{R})\cap L^1(\mathbb{R})$.. Given the wavelets $\psi(\cdot)$ and $\phi(\cdot)$, we construct wavelet sequences through translations and dilatations of mother and father wavelets, respectively, given by

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t-k\right),$$
$$\phi_{j,k}(t) = 2^{-j/2}\phi\left(2^{-j}t-k\right).$$

The functions $\{\psi_{j,k}(\cdot), j,k\in\mathbb{Z}\}$ and $\{\phi_{j,k}(\cdot), j,k\in\mathbb{Z}\}$ form bases that are not necessarily orthogonal. The advantage of working with orthogonal bases is that they allow the perfect reconstruction of a signal from the coefficients of the transform. In general, the most used orthogonal wavelets are: Haar, Daublets, Symmlets and Coiflets.

The fourth step consists in transform the observations $F(l)$, $l\in\{4,5,\cdots,g(n)\}$, into the symmlet wavelet "s8" domain by applying a discrete wavelet transform (see Definition 2.1), with level $J=\lfloor \log_2(g(n)-3)\rfloor$, to obtain a sequence of wavelet coefficients $\boldsymbol{d}_4, \boldsymbol{d}_5, \cdots, \boldsymbol{d}_J$.

Definition 2.1. (**Wavelet Transform**) Let $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)'$ be an i.i.d. random sample, with $J=\lfloor \log_2(n)\rfloor$, where $\lfloor\cdot\rfloor$ indicates the integer part function. The discrete wavelet transform (DWT) of $\boldsymbol{X}$, with respect to the mother wavelet $\psi(\cdot)$, is defined as

$$d_{j,k} = \sum_{t=1}^{n} X_t \psi_{j,k}(t), \tag{2.3}$$

for all $j=1,2,\cdots,J$ and $k=1,2,\cdots,\lfloor \frac{n}{2^j} \rfloor$. We can write the transform (2.3) in matrix form by

$$\boldsymbol{d}_j = \boldsymbol{W}_j \boldsymbol{X}, \tag{2.4}$$

where $\boldsymbol{W}_j=(\psi_{j,k}(t))_{k,t}$ is a $\lfloor \frac{n}{2^j} \rfloor \times n$ matrix. Assuming appropriate boundary conditions, the transform is orthogonal and one can obtain the *inverse discrete wavelet transform* (IDWT) given by

$$\boldsymbol{X} = \boldsymbol{W}'\boldsymbol{d},$$

where $\boldsymbol{W}'$ denotes the transpose of $\boldsymbol{W}$.

Then shrink the wavelet coefficients towards zero, to obtain new detail coefficients $\tilde{d}_4 \equiv \delta^H_{-\lambda}(d_4),\cdots,\tilde{d}_J \equiv \delta^H_{-\lambda}(d_J)$, where $\lambda=\hat{\sigma}\sqrt{2\log(g(n)-3)}$, $\hat{\sigma}$ is the estimated level of noise given by

$$\hat{\sigma} = \frac{median\left\{|d_{J-1,k}| : 0 \le k < 2^J\right\}}{0.6745.}$$

and the $\delta^H_\lambda(x)$ is the *hard* (H) *shrinkage* function defined by

$$\delta^H_\lambda(x) = \begin{cases} 0, & \text{if } |x| \le \lambda \\ x, & \text{if } |x| > \lambda. \end{cases}$$

Finally, apply the inverse discrete wavelet transform, to get the *wavelet shrinkage estimator* $\hat{\mathrm{F}}(l)$ of $\mathrm{F}(l)$, for all $l\in\{4,5,\cdots,g(n)\}$.

Under such condition, the smoothed fluctuations can be characterized by a scaling exponent H, which is the slope of the line when one regresses $\ln(\hat{\mathrm{F}}(l))$ on $\ln(l)$,

$$\hat{\mathrm{F}}(l) \sim \varphi l^{\mathrm{H}}, \tag{2.5}$$

where

- $0<H<0.5$ indicates intermediate long-range dependence;
- $\mathrm{H}=0.5$ indicates short long-range dependence;
- $0.5<H<1$ indicates long-range dependence.

By taking the logarithm of the relationship in (2.5), we obtain $\ln(\hat{\mathrm{F}}(l)) \sim \ln(\varphi)+\mathrm{H}\ln(l)$. Then by the least squares method, the estimator for the exponent H is given by

$$\hat{\mathrm{H}} = \frac{\sum_{l=4}^{g(n)-3} (\ln(l) - \bar{\mathrm{x}}) \ln\left(\hat{\mathrm{F}}(l)\right)}{\sum_{l=4}^{g(n)-3} (\ln(l) - \bar{\mathrm{x}})^2}, \quad (2.6)$$

where $\bar{x} = \frac{1}{g(n)-3} \sum_{l=4}^{g(n)-3} \ln(l)$. The usual choice of $g(n)$ is $\lfloor \frac{n}{10} \rfloor$, where $\lfloor \cdot \rfloor$ indicates the integer part function.

Under some conditions, Theorem 2.1 determines an optimal choice for $g(n)$. Consequently, it supplies an asymptotic optimal choice for the number of regressors in the SDFA procedure.

Theorem 2.1. *Let $\{X_t\}_{t \in \mathbb{R}^+}$ be a fractional Gaussian noise process. Consider $\{X_t\}_{t=1}^{n}$ a time series from this process. Suppose that, for all* $l \in \{4, 5, \cdots, g(n)\}$, *the random variable $Z_1^l, Z_2^l, \cdots, Z_{\bar{n}}^l$, given by expression* (2.1), *are independent identically distributed, with distribution function $\mathcal{N}(0, \sigma_l^2)$. Suppose that* $\lim_{g(n), n \to \infty} \lfloor \frac{n}{g(n)} \rfloor = \infty$.. *Then, the $\hat{H}$* estimator, given by expression (2.6),

(a) is U.M.V.U. for the H parameter,
(b) has normal distribution with $\mathbb{E}(\hat{\mathrm{H}}) = \mathrm{H}$ and

$$\mathrm{Var}\left(\hat{\mathrm{H}}\right) = \frac{1}{2n \sum_{j=1}^{g(n)-3} (x_j - \bar{x})^2}, \quad (2.7)$$

where $x_j = \ln(j+3)$, for $j \in \{1, \cdots, g(n)-3\}$.

Proof. See Linhares [12]. □

The Corollary 2.1 presents de proposed $g(n)$.

Corollary 2.1. *Let $\{X_t\}_{t \in \mathbb{R}^+}$ be a fractional Gaussian noise process. Consider $\{X_t\}_{t=1}^{n}$ a time series obtained from this process. Suppose, for all* $l \in \{4, 5, \cdots, g(n)\}$, *the random variables $Z_1^l, Z_2^l, \cdots, Z_{\bar{n}}^l$ given in expression* (2.1) *are independent and identically distributed, with distribution function $\mathcal{N}(0, \sigma_l^2)$. Let $g(n) = (\ln(n))^2$. Then, $\hat{H}$ given by expression* (2.6),

(a) is a U.M.V.U. estimator for H,
(b) has Gaussian distribution with expected value and variance given, respectively, by

$$\mathbb{E}\left(\hat{\mathrm{H}}\right) = \mathrm{H} \quad \text{and} \quad \mathrm{Var}\left(\hat{\mathrm{H}}\right) = \frac{1}{2n \sum_{j=1}^{g(n)-3} (x_j - \bar{x})^2}, \quad (2.8)$$

where $x_j = \ln(j+3)$, for all , $j \in \{1, \ldots, g(n)-3\}$, $\bar{x} = (g(n)-3)^{-1} \sum_{j=1}^{g(n)-3} x_j$ and $g(n) = \lfloor (\ln(n))^2 \rfloor$.

Proof. Consider $g(n) = \lfloor (\ln(n))^2 \rfloor$. Then,

$$\left\lfloor \frac{n}{g(n)} \right\rfloor = \left\lfloor \frac{n}{(\ln(n))^2} \right\rfloor.$$

As $\lim_{g(n), n \to \infty} \lfloor \frac{n}{g(n)} \rfloor = \lim_{g(n), n \to \infty} \lfloor \frac{n}{(\ln(n))^2} \rfloor = \infty$, the result follows from Theorem 2.1. □

Table 3.1
Estimation results for FGN time series with $n \in \{1000, 5000\}$.

n		1000				5000			
H	**$g(n)$**	Mean	Bias	mse	Var	Mean	Bias	mse	Var
0.1	$\lfloor n/10 \rfloor$	0.3139	0.2139	0.0461	0.0003	0.1696	0.0696	0.0048	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.2138	0.1138	0.0139	0.0010	0.1444	0.0444	0.0020	0.0003
0.15	$\lfloor n/10 \rfloor$	0.3402	0.1902	0.0364	0.0003	0.2117	0.0617	0.0038	0.0004
	$\lfloor \ln(n)^2 \rfloor$	0.2384	0.0884	0.0092	0.0014	0.1882	0.0382	0.0015	0.0004
0.2	$\lfloor n/10 \rfloor$	0.3747	0.1747	0.0310	0.0005	0.2538	0.0538	0.0029	0.0004
	$\lfloor \ln(n)^2 \rfloor$	0.2956	0.0956	0.0121	0.0031	0.2312	0.0312	0.0010	0.0006
0.25	$\lfloor n/10 \rfloor$	0.4175	0.1675	0.0288	0.0008	0.3001	0.0501	0.0025	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.3330	0.0830	0.0094	0.0026	0.2793	0.0293	0.0009	0.0004
0.3	$\lfloor n/10 \rfloor$	0.4520	0.1520	0.0238	0.0008	0.3412	0.0412	0.0017	0.0006
	$\lfloor \ln(n)^2 \rfloor$	0.3664	0.0664	0.0097	0.0056	0.3205	0.0205	0.0004	0.0007
0.35	$\lfloor n/10 \rfloor$	0.490	0.1408	0.0209	0.0011	0.3864	0.0364	0.0014	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.4087	0.0587	0.0077	0.0045	0.3677	0.0177	0.0004	0.0001
0.4	$\lfloor n/10 \rfloor$	0.5345	0.1345	0.0191	0.0010	0.4370	0.0370	0.0015	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.4782	0.0782	0.0091	0.0031	0.4180	0.0180	0.0005	0.0002
0.45	$\lfloor n/10 \rfloor$	0.5702	0.1202	0.0154	0.0010	0.4828	0.0328	0.0011	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.5077	0.0577	0.0089	0.0058	0.4673	0.0173	0.0004	0.0001
0.5	$\lfloor n/10 \rfloor$	0.5939	0.0939	0.0095	0.0007	0.5280	0.0280	0.0009	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.5292	0.0292	0.0066	0.0061	0.5125	0.0125	0.0003	0.0001
0.55	$\lfloor n/10 \rfloor$	0.6612	0.1112	0.0149	0.0026	0.5757	0.0257	0.0008	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.6207	0.0707	0.0089	0.0041	0.5606	0.0106	0.0004	0.0003
0.6	$\lfloor n/10 \rfloor$	0.7002	0.1002	0.0116	0.0016	0.6244	0.0244	0.0007	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.6267	0.0267	0.0080	0.0076	0.6107	0.0107	0.0003	0.0002
0.65	$\lfloor n/10 \rfloor$	0.7457	0.0957	0.0114	0.0023	0.6803	0.0303	0.0012	0.0003
	$\lfloor \ln(n)^2 \rfloor$	0.7096	0.0596	0.0071	0.0037	0.6690	0.0190	0.0008	0.0004
0.7	$\lfloor n/10 \rfloor$	0.7863	0.0863	0.0097	0.0024	0.7246	0.0246	0.0009	0.0003
	$\lfloor \ln(n)^2 \rfloor$	0.7122	0.0122	0.0108	0.0113	0.7126	0.0126	0.0005	0.0004
0.75	$\lfloor n/10 \rfloor$	0.8465	0.0965	0.0105	0.0012	0.7744	0.0244	0.0007	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.8085	0.0585	0.0114	0.0084	0.7649	0.0149	0.0004	0.0002
0.8	$\lfloor n/10 \rfloor$	0.8707	0.0707	0.0077	0.0029	0.8171	0.0171	0.0005	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.8073	0.0073	0.0067	0.0070	0.8076	0.0076	0.0003	0.0003
0.85	$\lfloor n/10 \rfloor$	0.9312	0.0812	0.0085	0.0020	0.8615	0.0115	0.0003	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.8794	0.0294	0.0084	0.0079	0.8519	0.0019	0.0003	0.0004
0.9	$\lfloor n/10 \rfloor$	0.9886	0.0886	0.0105	0.0028	0.9122	0.0122	0.0003	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.9700	0.0700	0.0099	0.0053	0.9046	0.0046	0.0003	0.0003
0.95	$\lfloor n/10 \rfloor$	1.0441	0.0941	0.0106	0.0019	0.9675	0.0175	0.0006	0.0003
	$\lfloor \ln(n)^2 \rfloor$	1.0064	0.0564	0.0107	0.0080	0.9637	0.0137	0.0006	0.0005

3. Simulations

Let $\{B_H(t)\}_{\mathbb{R}^+}$ be a continuous Gaussian process starting at zero, with mean $\mu = 0$, stationary increments, variance given by $\mathbb{E}(B_H^2(t)) = t^{2H}$ and covariance given by $\mathbb{E}(B_H(s)B_H(t)) = \frac{1}{2}\{s^{2H} + t^{2H} - |s-t|^{2H}\}$, for all $t, s \in \mathbb{R}^+$. Then, $\{B_H(t)\}_{\mathbb{R}^+}$ is called a *fractional Brownian motion process.* The process $\{X_t\}_{t\in\mathbb{R}^+}$ is a *fractional Gaussian noise*, if $\{X_t\}_{t\in\mathbb{R}^+}$ is the increment of a fractional Brownian motion, namely $X_t = B_H(t+1) - B_H(t)$, for all $t \in \mathbb{R}^+$. The index H is the Hurst exponent (see [9]). When $0.5 < H < 1$, we say that the process displays the long-range dependence property.

To investigate the effect of $g(n)$ in the smoothed detrended fluctuation analysis, we generate time series from fractional Gaussian noise (FGN) models, with 200 replications and length $n \in \{1000, 5000, 10000, 15000\}$. For estimating the Hurst exponent H we consider the proposed value for $g(n) = \lfloor \ln(n)^2 \rfloor$ (see Corollary 2.1) and the usual choice $\frac{n}{10}$ given in the literature (see [12]), where $\lfloor \cdot \rfloor$ indicates the integer part function. For each $g(n)$, we calculate the empirical values of the mean, the bias, the mean square error (mse) and the variance (var) values. The experiment results are presented in Tables 3.1 and 3.2.

The usual choice of $g(n) = \lfloor n/10 \rfloor$ generates a good estimator for the fractional parameter, but we can observe in Tables 3.1 and 3.2 that $g(n) = \lfloor \ln(n)^2 \rfloor$ gives better results, obtaining smaller bias for all values of H, even when the value of n is small. When n is large enough, with this choice of $g(n) = \lfloor \ln(n)^2 \rfloor$, for computational purposes, it has the advantage of requiring a smaller number of regressors for estimating the H parameter.

Table 4.3
Estimation results by SDFA method with $g(n) = \ln(n)^2$, for the parameter H in RR time series of healthy and unhealthy records.

Healthy record in the *nsrdb*		Unhealthy record in the *mitdb*	
Record's number	$\hat{H}$	Record's number	$\hat{H}$
16786	0.9446	102	0.2748
16795	0.9374	104	0.2413
16483	0.9469	105	0.2806
16539	0.8713	106	0.4291
16773	0.9776	107	0.4137
16483	0.9469	109	0.5073
16539	0.8950	110	0.2176
16773	0.9776	111	0.5256
17052	0.8294	114	0.4614
17453	0.9103	116	0.2876

4. Application

Clinicians refer to normal activity of the heart as a regular sinus rhythm (see [14]); however, a normal RR times series, i.e., a series of time intervals between subsequent heart beats, classified by a specialist as normal, fluctuates in a complex manner. This behavior prompts the researches Makowiec et al. [14] to apply modern tools of statistical physics to uncover hidden dependencies, especially those manifested by long-range power-law correlations. Cardiac arrhythmia is defined as a change in electrical activity within the heart which manifests as irregular heartbeats. In extreme cases, arrhythmia can be induced by

Table 3.2
Estimation results for FGN time series with $n \in \{10000, 15000\}$.

n		10000				15000			
H	$g(n)$	Mean	Bias	mse	Var	Mean	Bias	mse	Var
0.1	$\lfloor n/10 \rfloor$	0.1413	0.0413	0.0017	0.0001	0.1336	0.0336	0.0011	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.1288	0.0288	0.0008	0.0001	0.1295	0.0295	0.0008	0.0001
0.15	$\lfloor n/10 \rfloor$	0.1882	0.0382	0.0014	0.0003	0.1789	0.0289	0.0008	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.1767	0.0267	0.0007	0.0004	0.1753	0.0253	0.0006	0.0002
0.2	$\lfloor n/10 \rfloor$	0.2341	0.0341	0.0012	0.0000	0.2261	0.0261	0.0007	0.0000
	$\lfloor \ln(n)^2 \rfloor$	0.2232	0.0232	0.0006	0.0000	0.2231	0.0231	0.0006	0.0000
0.25	$\lfloor n/10 \rfloor$	0.2781	0.0281	0.0008	0.0001	0.2720	0.0220	0.0005	0.0000
	$\lfloor \ln(n)^2 \rfloor$	0.2677	0.0177	0.0004	0.0001	0.2682	0.0182	0.0003	0.0000
0.3	$\lfloor n/10 \rfloor$	0.3273	0.0273	0.0008	0.0000	0.3172	0.0172	0.0003	0.0000
	$\lfloor \ln(n)^2 \rfloor$	0.3178	0.0178	0.0004	0.0001	0.3134	0.0134	0.0002	0.0001
0.35	$\lfloor n/10 \rfloor$	0.3725	0.0225	0.0006	0.0001	0.3676	0.0176	0.0004	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.3633	0.0133	0.0003	0.0001	0.3634	0.0134	0.0002	0.0001
0.4	$\lfloor n/10 \rfloor$	0.4183	0.0183	0.0004	0.0001	0.4177	0.0177	0.0004	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.4100	0.0100	0.0002	0.0001	0.4143	0.0143	0.0003	0.0001
0.45	$\lfloor n/10 \rfloor$	0.4642	0.0142	0.0003	0.0001	0.4630	0.0130	0.0002	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.4554	0.0054	0.0001	0.0001	0.4587	0.0087	0.0002	0.0001
0.5	$\lfloor n/10 \rfloor$	0.5162	0.0162	0.0005	0.0003	0.5100	0.0100	0.0003	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.5076	0.0076	0.0003	0.0003	0.5061	0.0061	0.0002	0.0001
0.55	$\lfloor n/10 \rfloor$	0.5587	0.0087	0.0003	0.0002	0.6125	0.0125	0.0002	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.5513	0.0013	0.0002	0.0002	0.6071	0.0071	0.0002	0.0001
0.6	$\lfloor n/10 \rfloor$	0.6074	0.0074	0.0002	0.0001	0.6618	0.0118	0.0002	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.6069	0.0069	0.0002	0.0001	0.6571	0.0071	0.0002	0.0002
0.7	$\lfloor n/10 \rfloor$	0.7100	0.0100	0.0003	0.0002	0.7070	0.0070	0.0003	0.0003
	$\lfloor \ln(n)^2 \rfloor$	0.7074	0.0074	0.0003	0.0002	0.7070	0.0070	0.0003	0.0002
0.75	$\lfloor n/10 \rfloor$	0.7606	0.0106	0.0002	0.0001	0.7561	0.0061	0.0001	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.7588	0.0088	0.0002	0.0002	0.7537	0.0037	0.0002	0.0001
0.8	$\lfloor n/10 \rfloor$	0.8138	0.0138	0.0004	0.0002	0.8031	0.0031	0.0001	0.0001
	$\lfloor \ln(n)^2 \rfloor$	0.8100	0.0100	0.0006	0.0003	0.8030	0.0030	0.0001	0.0001
0.85	$\lfloor n/10 \rfloor$	0.8606	0.0106	0.0004	0.0003	0.8490	0.0010	0.0002	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.8600	0.0100	0.0006	0.0004	0.8490	0.0010	0.0002	0.0001
0.9	$\lfloor n/10 \rfloor$	0.9112	0.0112	0.0003	0.0002	0.9050	0.0050	0.0002	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.9102	0.0102	0.0005	0.0003	0.9017	0.0017	0.0002	0.0002
0.95	$\lfloor n/10 \rfloor$	0.9595	0.0095	0.0003	0.0002	0.9541	0.0041	0.0002	0.0002
	$\lfloor \ln(n)^2 \rfloor$	0.9550	0.0050	0.0005	0.0003	0.9510	0.0010	0.0005	0.0002

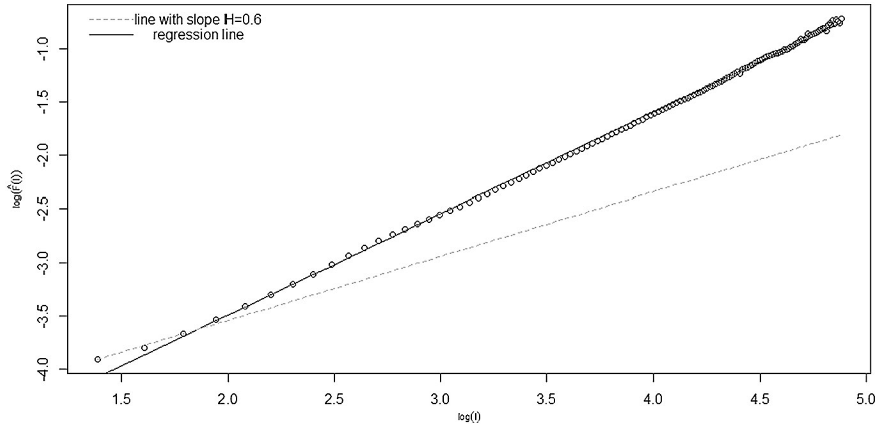

Fig. 1. Plots presenting the scaling properties of SDFA in log-scale for the healthy record in the *nsrdb* of number 16,786.

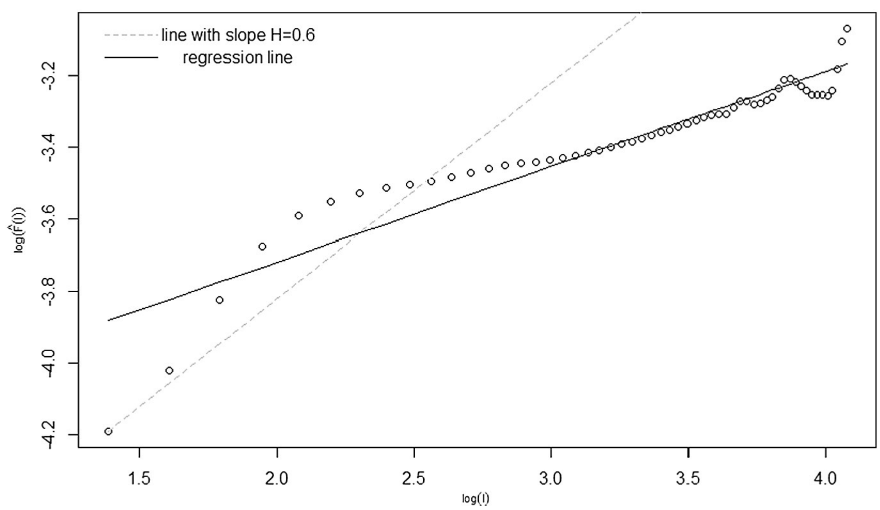

Fig. 2. Plots presenting the scaling properties of SDFA in log-scale for the unhealthy record in the *mitdb* of number 102.

damaged cardiac tissue or abnormal cardiac anatomy which may lead to a stroke or heart attack. Current methods to detect arrhythmias include electrocardiogram (ECG) readings and analysis. Many studies propose the Hurst estimators, as an additional diagnostic tool may provide an indication of cardiac arrhythmia (see [15]).

In this section in view of the SDFA method we analyze RR time series of 10 healthy and 10 unhealthy (with cardiac arrhythimia) randomly selected from databases available from PhysioBank (https://www.physionet.org/cgi-bin/atm/ATM) in order to check the difference between healthy and unhealthy RR time series. In the PhysionBank for the healthy RR time series we use the database named as *MIT-BIH Normal Sinus Rhythm Database* ***(nsrdb)*** and for uhealthy RR time series we use database named as *MIT-BIH Arrhythmia Database* ***(mitdb)***.

The Hurst parameter (H) is used to quantify long range dependence in time series data, here we consider the SDFA method (ver Section 2) to estimate the parameter H, considering the proposed $g(n) = \lfloor \ln(n)^2 \rfloor$ (ver Section 2.1). It is proposed that utilizing Hurst estimator by SDFA method, as an additional diagnostic tool may provide an indication of cardiac arrhythmia. Table 4.3 presents the Hurst estimates $\hat{H}$ by SDFA method with $g(n) = \ln(n)^2$, for the *healthy* and *unhealthy*(with cardiac arrhythimia) RR time series.

From Table 4.3 one observes that for *healthy* RR time series the estimates of $\hat{H}$ is always greater than 0.6. While for *unhealthy* RR times series the estimates of $\hat{H}$ is always less than 0.6. For each RR time series, this conclusion is statistically significant at 5% significance level. Therefore, there is a numerical evidence that for *healthy* RR time series we have $H \in (0, 0.6)$, while for *unhealthy* (with cardiac arrhythimia) RR times series the $H \in (0.6, 1)$.

Fig. 1 shows the fluctuation functions (in log-scale) of the SDFA method for the healthy Record in the *nsrdb* of number 16,786, where the slope of the regression line is bigger than $H = 0.6$. Fig. 2 shows the fluctuation functions (in log-scale) of the SDFA method for the unhealthy record in the *mitdb* of number 102 and we can note that the regression line is smaller than $H = 0.6$.

5. Conclusions

Under some conditions, we determine an optimal choice for the number of regressors in the SDFA method. In the literature, the usual choice of $g(n)$ is $\lfloor n/10 \rfloor$, but results reported here show that the proposed $g(n) = \lfloor \ln(n)^2 \rfloor$ is more efficient for determining the number of regressors in the *Smoothed Detrended Fluctuation Analysis* method. When n is large enough, with this choice of $g(n) = \lfloor \ln(n)^2 \rfloor$, for computational purposes, it has the advantage of requiring a smaller number of regressors for estimating the H parameter. It is proposed that utilizing Hurst estimator by SDFA method, as an additional diagnostic tool may provide an indication of cardiac arrhythmia. There is a numerical evidence that for *healthy* RR time series we have $H \in (0{,}0.6)$, while for *unhealthy* (with cardiac arrhythimia) RR times series the $H \in (0.6, 1)$. For each RR time series, this conclusion is statistically significant at 5% significance level.

References

[1] J.M. Bardet, I. Kammoun, "Asymptotic properties of the detrended fluctuation analysis of long-range dependence processes", IEEE Trans. Inf. Theory 54 (5) (2008) 2041–2052.

[2] N. Crato, R.R. Linhares, S.R.C. Lopes, "Statistical properties of detrended fluctuation analysis", J. Stat. Comput. Simul. 80 (6) (2010) 625–641.

[3] D.L. Donoho, I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", Biometrika 81 (1994) 425–455.

[4] D.L. Donoho, I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", J. Am. Stat. Assoc. 90 (1995) 1200–1224.

[5] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, "Wavelet shrinkage: asymptopia?", J. R. Stat. Soc. 57 (1995) 301–369.

[6] D. Grech, Z. Mazur, "Can one make any crash prediction in finance using the local Hurst exponent idea?", Physica A 336 (2004) 133–145.

[7] D. Grech, G. Pamula, "The local Hurst exponent of the financial time series in the vicinity of crashes on the Polish stock exchange market", Physica A 387 (2008) 4299–4308.

[8] D. Grech, "Global and local approaches describing critical phenomena on the developing and developed financial markets", Econophysics Approaches to Large-Scale Business Data and Financial Crisis, Springer, Tokyo 2010, pp. 149–172, http://dx.doi.org/10.1007/978–4–431-53853-0-7.

[9] H.R. Hurst, "Long-term storage in reservoirs", Trans. Am. Soc. Civ. Eng. 116 (1951) 770–799.

[10] E. Koscielny-Bunde, H.E. Roman, A. Bunde, S. Havlin, H.-J. Schellnhuber, Long-range power-law correlations in local daily temperature fluctuations, Phil. Mag. B 77 (1998) 1331–1340.

[11] A. Leite, A. Rocha, M. Silva, S. Gouveia, J. Carvalho, O. Costa, "Long-range dependence in heart rate variability data: ARFIMA modelling vs detrended fluctuation analysis DFA", Comput. Cardiol. (2007) 21–24, http://dx.doi.org/10.1109/CIC.2007.4745411.

[12] R.R. Linhares, Longa Dependência em sequências de DNA: análise de flutua ções destendenciadas, teorias das distribuiç ões estáveis e de waveletsPhD these Programa de Pós-Graduação em Matemática, Instituto de Matemática, UFRGS, Porto Alegre, 2011.

[13] R.R. Linhares, "Smoothed detrended fluctuation analysis", J. Stat. Comput. Simul. (2016), http://dx.doi.org/10.1080/00949655.2016.1154965.

[14] D. Makowiec., R. Gala, A. Dudkowska, A. Rynkiewicz, M. Zwierz, "Long-range dependencies in heart rate signals—revisited", Physica A: Stat. Mech. Appl. 369 (2) (2006) 632–644.

[15] K. Marwin, Applications of Long Range Dependence Characterization in Thermal Imagin e Heart Rate VariabilityPhD these University of California, 2015.

[16] C. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, "Mosaic organization of DNA nucleotides", Phys. Rev. E 49 (5) (1994) 1685–1689.

[17] C.K. Rhea, A.W. Kieferb, W.G. Wrighte, L.D. Raisbecka, F.J. Harang, "Interpretation of postural control may change due to data processing techniques", Gait Posture 41 (2) (2015) 731–735.

[18] S. Sammito, I. Böckelmann, "Factors influencing heart rate variability". Int. Cardiovasc. Forum J. **6** (2016) 18–22, http://dx.doi.org/10.17987/icfj.v6i0.242.

[19] L. Taizhi, Y.Q. Chen, M. Ko, B. Stark, "An online heart rate variability analysis method based on sliding window Hurst series", J. Fiber Bioeng. Inf. 8 (2) (2014) 391–400.

[20] M.S. Taqqu, V. Teverovsky, W. Willinger, "Estimators for long range dependence: an empirical study", Fractals 3 (4) (1995) 785–798.

[21] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, New York, 1999.

[22] R.-G. Yeh, J.-S. Shieh, Y.-Y. Han, Y.-J. Wang, S.-C. Tseng, "Detrended fluctuation analyses of short-term heart rate variability in surgical intensive care units", Biomed. Eng. Appl. Basis Commun. 18 (2006) 67–72.